

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number. 2001265539 A

(43) Date of publication of application: 28.09.01

(51) Int. Cl. G06F 3/06
G06F 12/16

(21) Application number. 2000074807

(22) Date of filing: 16.03.00

(71) Applicant. FUJI XEROX CO LTD

(72) Inventor. KYOZUKA SHINYA
HAMADA TSUTOMU
KAMIMURA TAKESHI
FUNADA MASAO

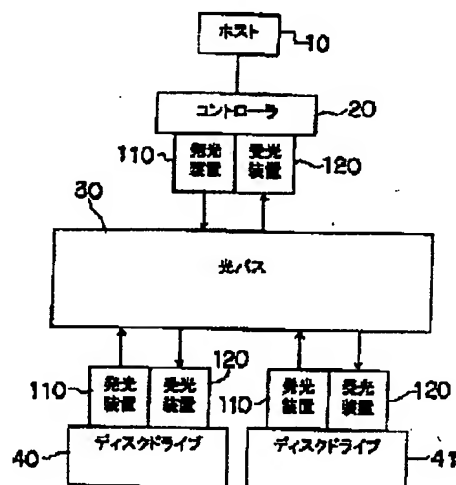
(54) ARRAY TYPE STORAGE DEVICE AND
INFORMATION PROCESSING SYSTEM

(57) Abstract

PROBLEM TO BE SOLVED: To enhance extensibility by suppressing lowering of data write speed in mirroring.

SOLUTION: A disk array controller 20 converts write data into an optical signal by a light emitting device 110 after storing the write data in a cache memory and transmits the optical signal to an optical bus 30. Signal light inputted in the optical bus 30 is broadcast transmitted by the optical bus 30 and simultaneously read in disk drives 40, 41 via a light receiving device. Thus, the lowering of data write speed in the mirroring is suppressed and the extensibility is enhanced.

COPYRIGHT: (C)2001,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-265539

(P2001-265539A)

(43) 公開日 平成13年9月28日 (2001.9.28)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード(参考)	
G 0 6 F 3/06	5 4 0	G 0 6 F 3/06	5 4 0	5 B 0 1 8
12/16	3 2 0	12/16	3 2 0 L	5 B 0 6 5

審査請求 未請求 請求項の数 4 O L (全 7 頁)

(21) 出願番号 特願2000-74807(P2000-74807)

(22) 出願日 平成12年3月16日(2000.3.16)

(71) 出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂二丁目17番22号

(72) 発明者 経塚 信也

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(72) 発明者 浜田 勉

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(74) 代理人 100079049

弁理士 中島 淳 (外3名)

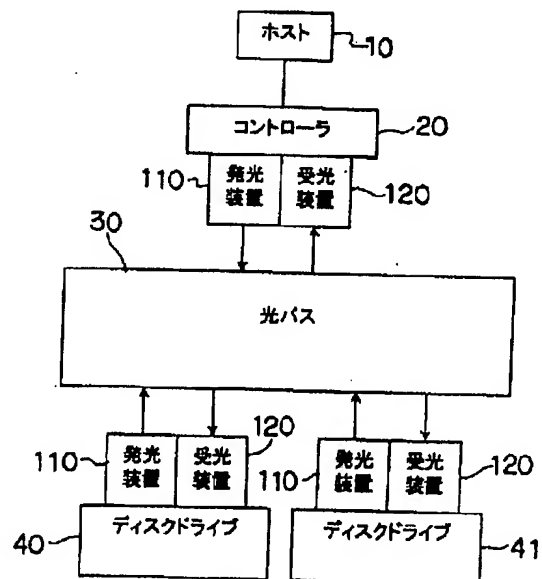
最終頁に続く

(54) 【発明の名称】 アレイ型記憶装置及び情報処理システム

(57) 【要約】

【課題】ミラーリングにおけるデータ書き込み速度の低下を抑制し、拡張性を高くする。

【解決手段】ホスト10からの書き込み要求に対して、ディスクアレイコントローラ20は書き込みデータをキャッシュメモリに格納した後、発光装置110により光信号に変換し、光バス30へ送出する。光バス30に入力された信号光は、光バス30でブロードキャスト伝送され、受光装置を介してディスクドライブ40、41に同時に読み込まれる。これにより、ミラーリングにおけるデータ書き込み速度の低下を抑制し、拡張性を高くすることができる。



【特許請求の範囲】

【請求項1】情報処理装置から入出力される情報が記憶されるアレイ型記憶装置であって、前記情報が記憶される複数の記憶装置と、前記情報処理装置及び前記複数の記憶装置の各々が接続され前記情報を入出力させる複数のアクセスポートを有し、前記情報が光信号として伝送されるブロードキャスト型光伝送媒体と、を備えたアレイ型記憶装置。

【請求項2】前記アクセスポートに接続され、前記複数の記憶装置のデータの入出力動作を制御する信号を前記他のアクセスポートを介して前記ブロードキャスト型光伝送媒体に出力し前記複数の記憶手段に伝送するコントローラを備えたことを特徴とする請求項1記載のアレイ型記憶装置。

【請求項3】前記コントローラは、前記複数の記憶装置をRAIDシステムとして動作するよう前記信号を出力することを特徴とする請求項2記載のアレイ型記憶装置。

【請求項4】データを入出力する情報処理装置と、前記データが記憶される複数の記憶装置と、前記情報処理装置及び前記複数の記憶装置の各々が接続され前記データを入出力させる複数のアクセスポートを有し、前記データが光信号として伝送されるブロードキャスト型光伝送媒体と、を備えたことを特徴とする情報処理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、コンピュータシステム等における2次記憶装置として用いられるアレイ型記憶装置に関し、特にデータを多重化して記録するアレイ型記憶装置、これを用いた情報処理システムに関する。

【0002】

【従来の技術】コンピュータシステムの2次記憶装置としては、一般に不揮発性記憶媒体が用いられ、代表的なものとして磁気ディスク装置、光ディスクなどがあげられる。近年、コンピュータシステムにおいて、この種の2次記憶装置に対してデータ転送速度や信頼性の向上などの高性能化が要求されており、これらの要求を実現する一つの方法として、多数の2次記憶装置(以下、ディスクドライブ)から構成されるアレイ型記憶装置が用いられている。

【0003】アレイ型記憶装置の構成として、RAID (Redundant Arrays of Independent Disks) がある。RAIDの構成としては、データをセクタまたはバイト単位で分割(ストライピング)してデータ転送を行う(RAID0)、複数のディスクドライブに全く同じデータを書き込むミラーリングを行う(RAID1)、ストライピングしたデー

タからパリティデータを生成して、データおよびパリティデータをそれぞれ専用のディスクドライブに格納する(RAID3、RAID4)、データおよびパリティデータを複数のディスクドライブに分散して格納する(RAID5)、がある。また、ミラーリングとパリティを組み合わせたRAID1+3、RAID1+4、RAID1+5といった構成もある。RAIDに関する公知文献としては、例えば、“The RAID Book: A STORAGE SYSTEM TECHNOLOGY HANDBOOK”, The RAID Advisory Board, 1997がある。

【0004】ここで、RAID1について簡単に説明しておく。1台のディスクアレイコントローラと2台のディスクドライブで構成されたRAID1においては、上位装置(ホスト)からのデータの書き込み要求に対しては、キャッシュメモリに書き込みデータを格納した後、2台のディスクドライブに対して同一データの書き込みを行う。ホストからのデータの読み出し要求に対して、ディスクアレイコントローラはキャッシュメモリに要求データが存在する場合は、そのデータをホストに返す。キャッシュメモリに要求データが存在していないときは、ディスクアレイコントローラは2台のディスクドライブの一方にアクセスし、データを読み出しホストに返すとともに、キャッシュメモリに格納する。いずれか一方のディスクドライブに障害が発生した場合には、もう一方のディスクドライブに対してアクセスしデータを読み出す。このようにして、ディスクドライブの障害によるデータ喪失を防止することができる。

【0005】なお、ミラーリングを行うディスクドライブ数は2台に限らず、3台以上のディスクドライブに対してミラーリングを行えば、ディスクドライブの二重障害(2台のディスクドライブに同時に障害が発生する)に対してもデータ喪失を防止することができる。

【0006】上述した方法によって、ディスクドライブの障害に対するデータ喪失を防止することができるが、ディスクアレイコントローラに障害が発生した場合には、ホストからのデータの読み出し、書き込み要求に対処できない。これに対して、ディスクアレイコントローラも複数台を設けて、通常時には1台のディスクアレイコントローラでホストとのデータの送受、ディスクドライブへのデータの読み書きを制御し、ディスクアレイコントローラに障害が発生した場合には、別のディスクアレイコントローラによって、ホスト、ディスクドライブとのデータの送受を行うようにして、より信頼性を高めたアレイ型記憶装置も提案されている。

【0007】

【発明が解決しようとする課題】ディスクアレイコントローラとディスクドライブ間のデータ転送は、SCSI (small computer system interface)バスなどの電気バスによって行われている。電気バスでは、ディスクアレイコントローラとディスクドライブ間の接続において、

同時にアクセスできるディスクドライブただ一つである。図5は1台のディスクアレイコントローラと2台のディスクドライブをSCSIバスによって接続し、ミラーリングを行う場合のデータ書き込み動作のタイミングチャートである。ディスクアレイコントローラはディスクドライブAとの接続を確保し、データをディスクドライブAへと転送する。ディスクドライブAへのデータ転送が終了したら、ディスクドライブAとの接続を解放して、ディスクドライブBとの接続を確保し、ディスクドライブAに転送したデータと同一のデータをディスクドライブBへと転送する。このようにデータを書き込む場合、ディスクアレイコントローラからは時分割で同じデータD₁（または、データD₂）を2回送出するため、ミラーリングを行わない場合に比べると書き込み速度が約半分に低下する。ディスクドライブの二重障害に対する信頼性を確保するため、3台以上のディスクドライブにミラーリングを行う構成をとる場合には、同じデータを送出する回数が増えるため、さらに書き込み速度は低下する、という問題がある。

【0008】また、ディスクアレイコントローラも複数台を設けて、より信頼性を高めたアレイ型記憶装置を構成する場合、1台のディスクドライブに対して複数のディスクアレイコントローラを接続する必要がある、ディスクアレイコントローラの台数に応じた配線が必要となる。従って一旦構成されたアレイ型記憶装置に、ディスクアレイコントローラやディスクドライブを増設して、システムを拡張しようとする場合、インターコネクションを再構成する必要がありシステムの拡張性が低い、という問題もある。

【0009】本発明は、上記技術的課題を解決するためになされたものであり、ミラーリングにおけるデータ書き込み速度の低下を抑制でき、また、拡張性の高いアレイ型記憶装置および情報処理システムを提供することを目的とする。

【0010】

【課題を解決するための手段】上記目的を達成するため本発明のアレイ型記憶装置は、情報処理装置から入出力される情報が記憶されるアレイ型記憶装置であって、前記情報が記憶される複数の記憶装置と、前記情報処理装置及び前記複数の記憶装置の各々が接続され前記情報を入出力させる複数のアクセスポートを有し、前記情報が光信号として伝送されるブロードキャスト型光伝送媒体と、を備えたことを特徴とする。

【0011】また、本発明の情報処理システムは、データを入出力する情報処理装置と、前記データが記憶される複数の記憶装置と、前記情報処理装置及び前記複数の記憶装置の各々が接続され前記データを入出力させる複数のアクセスポートを有し、前記データが光信号として伝送されるブロードキャスト型光伝送媒体と、を備えたことを特徴とする。

【0012】前記情報処理装置及び前記複数の記憶装置の各々は、ブロードキャスト型光伝送媒体のアクセスポートに直接光接続してもよく、ブロードキャスト型光伝送媒体のアクセスポイントに発光装置及び受光装置を設けて、これらとは電氣的に接続することでこのアクセスポイントに接続してもよく、前記情報処理装置及び前記複数の記憶装置のいずれか一方を直接光接続し、他方を電氣的に接続してもよい。従って、ブロードキャスト型光伝送媒体には直接光信号を入出力することも、電気信号を介してブロードキャスト型光伝送媒体に変換された光信号を伝送することもできる。また、情報処理装置は、例えばコントローラなどを介してブロードキャスト型光伝送媒体に接続してもよい。即ち、伝送目的の情報がこのブロードキャスト型光伝送媒体中を光信号として伝送されるように情報処理装置及び複数の記憶装置が各々接続されればよい。

【0013】

【発明の実施の形態】以下、本発明の実施形態について説明する。

【0014】図1は本発明の第1の実施形態におけるアレイ型記憶装置を備えるシステムのブロック図であり、1台のディスクアレイコントローラと2台、すなわち複数台のディスクドライブとでRAID1を構成した場合である。図1において、10はホスト、20はディスクアレイコントローラ、30は光バス、40及び41はディスクドライブであり、ディスクアレイコントローラ20およびディスクドライブ40、41は電気信号を光信号に変換する発光装置110および光信号を電気信号に変換する受光装置120を備えている。ディスクドライブ40、41は、ハードディスク装置や光磁気(MO)記録装置、DVD-RAMなどの記録メディアが着脱可能な回転型記録装置、DATなどのテープ型記録装置、半導体メモリといった不揮発性記憶装置からなる。発光装置110及び受光装置120の各々は、光バス30のアクセスポートに光接続されている。ホスト10からの書き込み要求に対して、ディスクアレイコントローラ20は書き込みデータをキャッシュメモリ(図示せず)に格納した後、発光装置110により光信号に変換し、光バス30へ送出する。

【0015】光バス30としては、例えば特開平10-123350号公報に開示された光バスを用いることができる。この光バスはシート状の光伝送媒体を用い、入力された信号光を入力部に設けた光拡散手段によって拡散伝播させることにより、複数のポート間へのブロードキャスト伝送を可能としたものである。

【0016】光バス30に入力された信号光は、ブロードキャスト伝送されるため、ディスクドライブ40、41は同時に信号光を読み込むことができる。ディスクドライブ40、41では、送信された信号光を受光装置120により電気信号に変換し、データの書き込みを行

う。

【0017】図2は、本実施形態におけるデータ書き込み時のタイミングチャートを示したものであり、ディスクアレイドコントローラ20からデータを1回送出するだけで、2台のディスクドライブへのデータ書き込みが終了するため、ミラーリングによる書き込み速度の低下が生じない。

【0018】一方、ホスト10からのデータの読み出し要求に対しては、ディスクアレイドコントローラ20はキャッシュメモリに要求データが存在する場合は、そのデータをホスト10に返す。キャッシュメモリに要求データが存在していないときは、ディスクアレイドコントローラ20は2台のディスクドライブ40、41の一方、例えばディスクドライブ40に対してのデータの読み出し命令を含んだ光信号を光バスへ出力する。ディスクドライブ40のデータは、発光装置110により光信号に変換され、光バス30を介してディスクアレイドコントローラ20の受光装置120へと伝送され、受光装置120によって電気信号に変換される。ディスクアレイドコントローラ20は、このデータをホスト10に返すとともに、キャッシュメモリに格納する。ディスクドライブ40に障害が発生した場合には、もう一方のディスクドライブ41から同様な手順でデータを読み出すことができる。

【0019】図3として、図1に示したアレイド型記憶装置に対して、新たにディスクアレイドコントローラ21及びディスクドライブ42を追加し、コントローラの障害に対する信頼性を高めるようにした例を示す。光バス30の他のアクセスポイントの各々に対してディスクアレイドコントローラ21、ディスクドライブ42を接続し、ディスクアレイドコントローラ20とディスクアレイドコントローラ21間だけを新たにバスケーブル50で接続してある。2台のディスクアレイドコントローラ間、すなわち複数のディスクアレイドコントローラ間では、バスケーブル50を介して互いのキャッシュメモリのデータを等しくする(キャッシュミラーリング)よう動作する。ディスクアレイドコントローラの一方は、ホストから選択的に読み出し書き出しの指示がなされたのを受けて、該当するディスクドライブを特定する信号とともに動作指令信号をアクセスポートを介して光伝送する。指定されたディスクドライブは、光信号を受信すると、該当の動作を行い、データをアクセスポートから受信あるいは出力する。ホストはディスクアレイドコントローラの一方が不調の場合には、動作させるディスクアレイドコントローラを変更する動作を行う。

【0020】なお、複数のディスクアレイドコントローラをコントロールするコントローラをさらに上位に設け、キャッシュミラーリングやディスクアレイドコントローラを制御させることでバスケーブルを省略することも可能である。上位コントローラとディスクアレイドコントロー

ラを接続する電気配線をそれぞれ設ける場合にはディスクコントローラの増設拡張性が劣るが、ディスクコントローラ自体は2乃至3系統あれば十分に信頼性を高められる場合が多いので、ディスク増設の拡張性が主として要求される場合にはこの形態も十分有効である。

【0021】このように、ブロードキャスト型光伝送媒体を用いることで、すべてのインターコネクションを再構成する必要がなく、容易にシステムの拡張が可能となる。

【0022】なお、本実施形態においては、光バス30として、シート状の光伝送媒体を用い、入力された信号光を光拡散手段によって拡散伝播させる光バスを用いているが、これに限定されるものではなく、光信号を1つのアクセスポートから並行して複数のアクセスポートに伝送できるブロードキャスト伝送型光伝送媒体であればよく、さらにこの伝送媒体が記憶装置とコントローラ間を光信号の双方向伝送可能な媒体であれば、よりシステムの簡略化と拡張性が向上する。また、拡散伝播型のほかに例えば光スターカブラや、光分岐器、光結合器などを組み合わせて、用いることができる。

【0023】また、発光装置110及び受光装置120が、ディスクアレイドコントローラ20、ディスクドライブ40、41(図3ではディスクアレイドコントローラ20、21、ディスクドライブ40、41、42)に設けてあるが、光バス30に発光装置110及び受光装置120を設けて、ディスクアレイドコントローラ20、ディスクドライブ40、41(図3ではディスクアレイドコントローラ20、21、ディスクドライブ40、41、42)との接続インターフェースを電気信号としても構わない。すなわち、本発明における光バスとは、光信号をブロードキャスト伝送を行う機能を有するものであって、その接続インターフェースは光信号でも、電気信号でも、あるいはこれらが混在していてもよい。光バスにアクセスポイントになる発光装置及び受光装置を設けた場合には、ディスクアレイドコントローラ及びディスクドライブは、光バスに電気的に接続されることになる。

【0024】なお、ホスト、光バス、記憶装置あるいはコントローラは、全てあるいはいくつか切り離し可能として構成してもよい。このときコントローラとホストとは直接接続するよう構成する方が、伝送速度の面で有利である。また、接続される記憶装置側に光電変換部を備えていない場合には、コネクタまでは電気接続とし、電気信号を光電変換する装置を設けることもできる。この場合、光電変換装置を駆動するための電源はホスト側からコントローラを介して供給するようにしても良いし、直接電源に接続することも可能である。

【0025】図4は本発明の第2の実施形態におけるアレイド型記憶装置のブロック図であり、2台、すなわち複数のディスクアレイドコントローラと6台、すなわち複数のディスクドライブでRAID1+4を構成した場

合である。図4において、10はホスト、20、21はディスクアレイドコントローラ、30、31は光バス、40から45はディスクドライブであり、光バス31、32は光伝送路100と発光装置110、受光装置120から構成される。ディスクドライブ40、41、42、およびディスクドライブ43、44、45をそれぞれ一つのブロック単位としてRAID4を構成しており、ディスクドライブ40、41、43、44はデータ格納用、ディスクドライブ42、45はパリティ格納用である。2つ、すなわち複数のブロック間でミラーリングを行う。各ブロックは、データ格納用、パリティ格納用の機能が異なる複数種類のディスクドライブを備えている。複数台のディスクアレイドコントローラ、複数種類のディスクドライブを備えた複数のディスクドライブブロックは、光バス31のアクセスポイントに光接続されている。

【0026】ホスト10からの書き込み要求に対して、ホスト10からのデータは光バス30によってディスクアレイドコントローラ20、21へと同時に伝送される。ディスクアレイドコントローラ20、21は、各々書き込みデータをキャッシュメモリに格納する。いずれか一方のディスクアレイドコントローラ、例えばディスクアレイドコントローラ21は、ホスト10から転送されたデータを、ストライピング回路、パリティ生成回路(いずれも、図示せず)によって、ストライピングおよびパリティ生成を行い、データおよびパリティデータを時分割して光バス31へ送出する。光バス31より送出されたデータおよびパリティデータは、ミラーリングを行う2つのブロックに同時に転送される。例えば、データをD1、D2、パリティデータPとすると、ディスクアレイドコントローラ21は、まずデータD1を送出し、データ格納用ディスクドライブ40、43へと書き込む。次に、データD2を送出し、データ格納用ディスクドライブ41、44へと書き込む。最後に、パリティデータPを送出し、パリティ格納用ディスクドライブ42、45へと書き込む。このように各ブロック内ディスクドライブへのデータ書き込みは時分割で行われるが、同一データの2台のディスクドライブへの書き込みは同時に行われるため、ミラーリングによる書き込み速度の低下は生じない。

【0027】一方、ホスト10からのデータの読み出し要求に対しては、ディスクアレイドコントローラ20、21のいずれかは、キャッシュメモリに要求データが存在する場合は、そのデータをホスト10に返す。キャッシュメモリに要求データが存在していないときは、いずれか一方のディスクアレイドコントローラ、例えばディスクアレイドコントローラ20は、2つのディスクドライブのブロックいずれか一方、例えばディスクドライブ40、41、42に対しての読み出し要求を含んだ光信号を光バスに入力する。データ格納用ディスクドライブ40、

41に障害が発生していない場合には、データ格納用ディスクドライブ40、41から読み出し要求データを光バス31へ送出する。読み出し要求データがデータ格納用ディスクドライブ40、41にまたがって存在している場合には、書き込み要求の場合と同様にして、データを時分割で光バス31へと送出する。データ格納用ディスクドライブ40、41のいずれかに障害が発生している場合には、障害のないデータ格納用ディスクドライブとパリティ格納用ディスクドライブ42からデータおよびパリティデータを光バス31へ送出する。ディスクドライブ42、41、42から送出されたデータおよびパリティデータは光バス31によって、2台のディスクアレイドコントローラ20、21へと同時に転送される。2台のディスクアレイドコントローラ20、21では、データ格納用ディスクドライブ40、41に障害が発生していない場合には、データを各々のキャッシュメモリに格納し、いずれかのデータ格納用ディスクドライブに障害が生じている場合には、データとパリティデータからデータを生成して、各々のキャッシュメモリに格納し、一方のディスクアレイドコントローラからホスト10へとデータを返す。

【0028】本実施形態に示したようなディスクアレイドコントローラが複数ある構成では、前述したように各ディスクアレイドコントローラ20、21のキャッシュメモリのデータを等しくすること(キャッシュミラーリング)が必要である。上述したように、ホスト10と2台のディスクアレイドコントローラ20、21が光バス30で接続されているため、ホスト10からの書き込みデータは同時に2台のディスクアレイドコントローラ20、21へと転送される。また、ディスクドライブから読み出しデータも、ディスクアレイドコントローラとディスクドライブが光バス31で接続されているため、2台のディスクアレイドコントローラ20、21へと転送される。従って、前述したように、ディスクアレイドコントローラ20とディスクアレイドコントローラ21間を接続するバスを設けなくてもキャッシュミラーリングができる。また、ホスト10とディスクアレイドコントローラ20、21間を光バス30によって接続することによって、ディスクアレイドコントローラの増設に対して、新たにディスクアレイドコントローラ間をバスケーブルで接続する必要がないという利点もある。

【0029】本発明は、前述のRAID方式にかかわらず適用することが可能である。

【0030】RAID0システムに適用する場合には、拡張性を有すること以外に、電気配線を用いた従来のストライピングに比べ、高速化することもできる。即ち、RAID0システムにおいては、時分割で複数のディスクドライブにデータの書き込み、読み出しを行わせるので、データ転送速度は、コントローラI/O>コントローラとディスクドライブ間のバス>ディスクI/Oの順でな

* て、十分早いので、従来の電気バスで並列に接続した場合と同等のデータ転送速度をえることができる。また、波長多重、強度多重といった多重化技術を用いれば、複数の異なるディスクブロックに対して同時にアクセスすることができ、並列動作が可能となり、さらにデータ転送速度を高めることができる。

【００３４】
【発明の効果】以上説明したように、本発明によれば、ミラーリングにおけるデータ書き込み速度の低下を抑制でき、また、拡張性が高いアレイ型記憶装置及び情報処理システムを得ることができる、という効果が得られる。

【図１】本発明の第１の実施形態を示すアレイ型記憶装置を備えたシステムのブロック図である。

【図3】第1の実施形態の変形例を示すアレイ型記憶装置のブロック図である。

【図４】第２の実施形態を示すアレイ型記憶装置のブロック図である。

【図5】従来のアレイ型記憶装置の、データ書き込み時のタイミングチャートを示す図である。

10…ホスト

20. 21…ディスクアレイコントローラ

30、31…光バス

40、41、42、43、44、45…ディスクドライ
ブ

50…バスケーブル

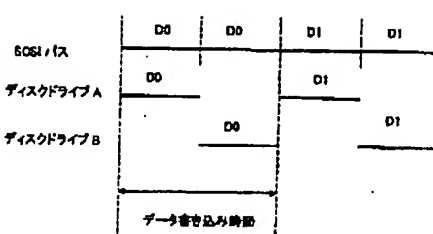
100...光伝送路

110…發光裝置

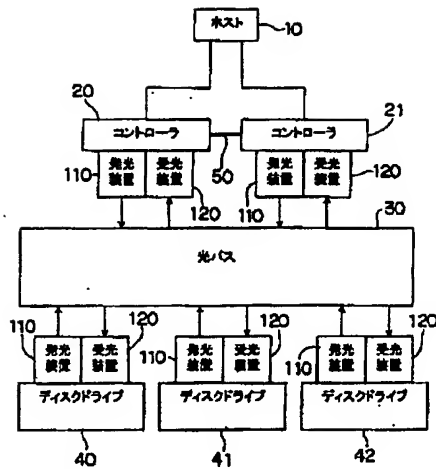
120…受光装置

のデータ転送速度、コントローラI/Oの転送速度に比べ *

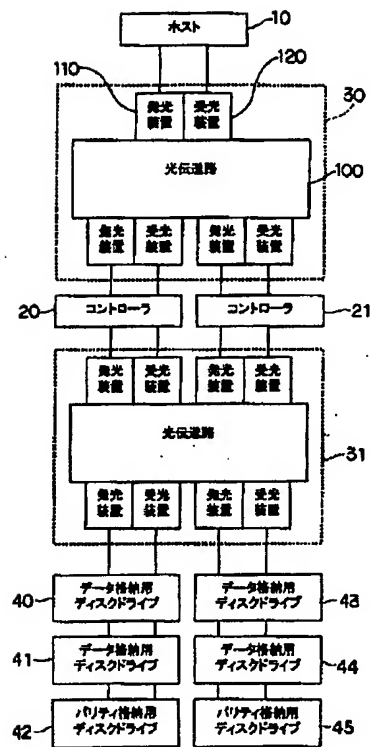
【圖5】



【図3】



【図4】



フロントページの続き

(72)発明者 上村 健
神奈川県足柄上郡中井町境430 グリーン
テクなかい 富士ゼロックス株式会社内

(72)発明者 舟田 雅夫
神奈川県足柄上郡中井町境430 グリーン
テクなかい 富士ゼロックス株式会社内
Fターム(参考) 5B018 GA04 HA04 MA14
5B065 BA01 CA18 CA30